



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12506

To cite this version : Durix, Bastien and Quéau, Yvain and Charvillat, Vincent and Durou, Jean-Denis [*Quels prétraitements pour la stéréophotométrie non calibrée ?*](#) (2013) In: Journées francophones des jeunes chercheurs en vision par ordinateur - ORASIS 2013, 10 June 2013 - 14 June 2013 (Cluny, France).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Quels prétraitements pour la stéréophotométrie non calibrée ?

Bastien Durix

Yvain Quéau

Vincent Charvillat

Jean-Denis Durou

IRIT, UMR CNRS 5505, Toulouse

bastien.durix@enseeiht.fr yvain.queau@enseeiht.fr charvi@enseeiht.fr durou@irit.fr

Résumé

La stéréophotométrie est une technique de scannage 3D qui consiste à prendre plusieurs photographies d'un objet sous le même angle, mais sous différents éclairages. Lorsque les éclairages sont inconnus, le problème se complique et sa résolution nécessite de faire plusieurs hypothèses, dont deux sont communes à toutes les méthodes existantes : l'hypothèse lambertienne et la contrainte d'intégrabilité. Nous proposons d'effectuer des prétraitements sur les images, afin de limiter l'influence d'éventuels écarts à ces hypothèses. Nous préconisons pour ce faire de retirer les pixels suspects. Nous validons ces améliorations sur un jeu d'images réelles de visages accessibles sur Internet.

Mots Clef

Stéréophotométrie, loi de Lambert, intégrabilité.

Abstract

Photometric stereo is a 3D-scanning technique which consists in taking several photographs of an object under the same angle but under different lightings. When the lightings are unknown, the problem is more complex. To solve it, further assumptions are required, amongst which the Lambertian assumption and the constraint of integrability. We show that some well-chosen preprocessings on the images may limit the influence of possible deviations from those assumptions. We suggest that the suspected pixels should be discarded. We validate these improvements on a database of real images of faces available on the web.

Keywords

Photometric stereo, Lambert's law, integrability.

1 Introduction

La stéréophotométrie, technique imaginée par Woodham au début des années 1980 [19], connaît actuellement un grand succès. Contrairement aux autres techniques de reconstruction 3D, elle permet d'estimer la couleur (ou *albédo*) des objets, et pas seulement leur relief. Son principe consiste à prendre plusieurs photographies d'un objet sous le même point de vue, mais sous différents éclairages. Lorsque les conditions d'éclairage sont connues, il s'agit de *stéréophotométrie calibrée* (CPS, pour *Calibrated Photometric Stereo*). Sous l'hypothèse lambertienne, on peut

alors très simplement estimer le champ de normales et l'albédo en chaque pixel. Pour obtenir le relief, il reste ensuite à *intégrer* le champ de normales [6].

Lorsque les conditions d'éclairage sont inconnues, on parle de *stéréophotométrie non calibrée* (UPS, pour *Uncalibrated Photometric Stereo*). Il s'agit d'un problème mal posé car, en plus des normales et de l'albédo, les éclairages doivent également être estimés. Pour rendre le problème UPS bien posé, une première hypothèse, connue sous le nom de *contrainte d'intégrabilité*, est ajoutée à l'hypothèse lambertienne. Dans la version calibrée de la stéréophotométrie, cette hypothèse est également utilisée, mais seulement dans l'étape d'intégration du champ de normales.

Nous verrons dans le paragraphe 2 que le problème UPS comporte une ambiguïté résiduelle connue sous le nom d'*ambiguïté de bas-relief* [1]. Une hypothèse supplémentaire est donc requise pour sa résolution. Dans cet article, nous allons supposer que l'éclairage a la même intensité dans les différentes prises de vue, mais des hypothèses très différentes ont été proposées, par exemple que l'objet photographié est constitué de régions ayant une couleur homogène (comme un costume d'Arlequin). On comprend, dès lors, pourquoi il est délicat de comparer des méthodes de résolution de l'UPS reposant sur des hypothèses aussi disparates.

Notre travail vise à améliorer les performances d'une méthode de résolution de l'UPS grâce à des prétraitements judicieux. De telles améliorations pourraient être utiles à toutes les méthodes de résolution de l'UPS, car elles s'attachent à garantir au mieux leurs deux hypothèses communes (hypothèse lambertienne et contrainte d'intégrabilité), qui ne sont jamais vérifiées en pratique. Nous exposons dans le paragraphe 2 les principales étapes de la résolution des problèmes CPS et UPS. Nous faisons ensuite, dans le paragraphe 3, une analyse critique de la PS, en détaillant l'ensemble des hypothèses sur lesquelles repose cette technique. La contrainte d'intégrabilité et l'hypothèse lambertienne sont explorées en profondeur dans les paragraphes 4 et 5, respectivement. Le paragraphe 6, qui est dédié à des tests sur images réelles, vise à valider ce travail en montrant que la résolution du problème UPS peut grandement bénéficier des prétraitements suggérés. Le paragraphe 7 résume nos principales contributions et ouvre quelques perspectives.

2 Résolution de la stéréophotométrie

2.1 Stéréophotométrie calibrée

La *shape-from-shading* est une technique de reconstruction 3D connue pour être mal posée [13]. Woodham montre dans [19] que l'utilisation de m photographies d'un objet prises sous le même point de vue, mais sous m éclairages différents, rend le problème bien posé, pour peu que les m « cartes de réflectance » (*reflectance maps*) qui leur correspondent soient connues [12]. Cette technique s'appelle la *stéréophotométrie calibrée*. De telles cartes dépendent non seulement de la BRDF (*Bidirectional Reflectance Distribution Function*) du matériau, mais aussi de l'éclairage, donc il n'est pas du tout évident d'en disposer en pratique. C'est pourquoi on suppose généralement que les éclairages sont directionnels et que le matériau est *lambertien*. L'éclairage de la $i^{\text{ème}}$ prise de vue peut alors être décrit, en direction et en norme, par un vecteur $S^i = [S_x^i, S_y^i, S_z^i]^\top$ orienté dans le sens opposé à la lumière, et le niveau de gris du pixel p dans la $i^{\text{ème}}$ image s'écrit alors simplement :

$$I_p^i = \rho_p N_p S^i \quad (1)$$

où ρ_p désigne l'albédo au pixel p et $N_p = [N_x, N_y, N_z]^\top(p)$ le vecteur normé normal à la surface en p . En notant $I_p = [I_p^1, \dots, I_p^m]$, $S = [S^1, \dots, S^m]$ et $M_p = \rho_p N_p$, on tire de (1) le système d'équations linéaires suivant :

$$I_p = M_p S \quad (2)$$

Si la matrice S est de rang 3, ce qui signifie que les m vecteurs d'éclairage S^i ne sont pas tous coplanaires, la solution aux moindres carrés de (2) est $\widehat{M}_p = I_p S^+$, de laquelle on tire l'albédo et la normale en p :

$$\widehat{\rho}_p = \|\widehat{M}_p\| \quad \widehat{N}_p = \frac{\widehat{M}_p}{\|\widehat{M}_p\|} \quad (3)$$

Il est également possible d'obtenir simultanément l'ensemble des albédos et des normales. Pour ce faire, on empile les vecteurs I_p correspondant à tous les pixels p situés à l'intérieur du masque Ω de l'objet, pour former une matrice $I = [I_1^\top, \dots, I_{|\Omega|}^\top]^\top$, où $|\Omega|$ désigne le cardinal de Ω . Si on construit de même $M = [M_1^\top, \dots, M_{|\Omega|}^\top]^\top$, alors :

$$I = M S \quad (4)$$

d'où l'on tire :

$$\widehat{M} = I S^+ \quad (5)$$

On peut évaluer la précision de la reconstruction 3D en calculant l'écart angulaire moyen (EAM) entre la vérité terrain et les normales \widehat{N} obtenues par normalisation de \widehat{M} .

2.2 Stéréophotométrie non calibrée

Lorsque la matrice S est inconnue, le problème est mal posé. Par décomposition en valeurs singulières (SVD), la matrice I se factorise en :

$$I = U W V^\top \quad (6)$$

avec $U \in \mathbb{R}^{|\Omega| \times |\Omega|}$, $W \in \mathbb{R}^{|\Omega| \times m}$ et $V \in \mathbb{R}^{m \times m}$. Les éclairages et les normales étant des vecteurs de \mathbb{R}^3 , le rang de I devrait être égal à 3, donc seules ses 3 plus grandes valeurs singulières devraient être non nulles. Cela n'arrive jamais en pratique, à cause du bruit et de divers écarts au modèle que nous détaillerons plus loin, mais il est légitime de restreindre W à sa sous-matrice principale \overline{W} de dimension 3, et U et V à leurs 3 premières colonnes, c'est-à-dire d'introduire $\overline{U} \in \mathbb{R}^{|\Omega| \times 3}$, $\overline{W} \in \mathbb{R}^{3 \times 3}$ et $\overline{V} \in \mathbb{R}^{m \times 3}$ tels que :

$$I \approx \overline{I} = \overline{U} \overline{W} \overline{V}^\top \quad (7)$$

où \overline{I} est une matrice proche de I qui, par construction, est de rang 3. La solution du problème $\overline{I} = M S$ s'écrit alors :

$$\begin{cases} \widehat{M} = \overline{U} P^\top \\ \widehat{S} = Q \overline{V}^\top \end{cases} \quad (8)$$

où P et Q sont deux matrices 3×3 , a priori quelconques, telles que :

$$P^\top Q = \overline{W} \quad (9)$$

Néanmoins, le problème est mal posé puisqu'il existe une infinité de tels couples (P, Q) . Yuille et Snow montrent dans [21] que la *contrainte d'intégrabilité* :

$$\frac{\partial}{\partial y} \frac{N_x}{N_z} - \frac{\partial}{\partial x} \frac{N_y}{N_z} = 0 \quad (10)$$

permet d'estimer six des neuf coefficients de P^{-1} :

$$P^{-1} = \begin{pmatrix} a_1 & a_2 & b_1 \\ a_3 & a_4 & b_2 \\ a_5 & a_6 & b_3 \end{pmatrix} \quad (11)$$

en l'occurrence les coefficients a_i , $i = 1 \dots 6$. Les trois autres coefficients (b_1, b_2, b_3) restent inconnus, ce qui correspond à une ambiguïté résiduelle appelée *ambiguïté de bas-relief* [1].

Pour lever l'ambiguïté de bas-relief, plusieurs approches ont été proposées, qui reposent sur des hypothèses très diverses. Nous supposons que les éclairages ont tous la même intensité, c'est-à-dire que les vecteurs S^i , $i \in [1, m]$, sont tous de même norme S_0 inconnue. La méthode que nous utilisons est décrite dans [17]. Elle consiste à résoudre le problème d'optimisation suivant :

$$(\widehat{b}, \widehat{S}_0) = \underset{(b, S_0) \in \mathbb{R}^3 \times \mathbb{R}^+}{\operatorname{argmin}} \|A b - K(S_0)\|^2 \quad (12)$$

où $b = [b_1, b_2, b_3]^\top$, $A = \overline{V} \overline{W}$ et :

$$K(S_0) = \begin{pmatrix} \sqrt{S_0^2 - S_x^2 - S_y^2} \\ \vdots \\ \sqrt{S_0^2 - S_x^2 - S_y^2} \end{pmatrix} \quad (13)$$

La solution de (12) en b permet de terminer l'estimation (11) de P^{-1} , puis celle de Q grâce à (9), dont on peut finalement déduire les estimations (8) de M et de S .

La précision de la reconstruction 3D et celle du calibrage des éclairages sont évaluées par l'EAM entre la vérité terrain et les vecteurs \widehat{M} et \widehat{S} normalisés.

3 Stéréophotométrie : étude critique

3.1 Avantages de la stéréophotométrie

Au vu du paragraphe précédent, la stéréophotométrie (PS) semble très simple à mettre en œuvre, que ce soit dans le cas calibré (CPS) ou dans le cas non calibré (UPS). Elle présente effectivement plusieurs avantages significatifs sur la stéréoscopie. En particulier, elle permet d'estimer simultanément le relief et l'albédo de l'objet photographié.

Un autre avantage vient de ce que l'appareil photographique prend toutes les vues sous le même angle. Il n'est donc pas nécessaire d'apparier les points, car un point physique correspond au même pixel dans les différentes images. Il s'ensuit que la résolution de la PS est bien plus rapide que celle de la stéréoscopie. En outre, comme le montrent les équations (5) et (8), le calcul de la solution est élémentaire et peut être mené de front pour tous les pixels. Un problème fréquent en stéréoscopie est que certains points ne peuvent pas être appariés, à cause soit d'une occultation, soit d'un manque de texture sur certaines parties de la surface. Il en résulte que les cartes de profondeur obtenues en stéréoscopie peuvent comporter des trous, ce qui ne peut arriver en PS.

Enfin, on peut mettre en œuvre la PS avec des images provenant d'un appareil photographique non calibré, chose inconcevable en stéréoscopie. D'une part, aucun calibrage externe n'est requis, puisque l'appareil photographique est fixe. D'autre part, nous n'avons à aucun moment précisé si le modèle de projection était perspectif ou non, ce qui semble indiquer que le calibrage interne non plus n'est pas requis. Effectivement, le choix d'un modèle de projection n'a pas d'influence sur l'estimation de l'albédo et des normales. En revanche, la connaissance de la matrice de projection est quand même nécessaire pour calculer le relief à partir des normales estimées [6].

3.2 Difficultés de la stéréophotométrie

Si la PS présente autant d'avantages vis-à-vis de la stéréoscopie, comment se fait-il que cette dernière reste encore aujourd'hui la technique de référence en matière de reconstruction 3D photographique ? Primo, la PS ne peut être mise en œuvre qu'en conditions très contrôlées : la scène doit être rigide et immobile, et l'appareil photographique doit être posé sur un pied. Il a bien été montré dans [10] que l'utilisation de $m = 3$ sources de lumière colorées (en rouge, vert et bleu) permettait de prendre $m = 3$ photographies simultanées d'une scène déformable. Néanmoins, cette technique reste surtout une technique de laboratoire. Secundo, l'intégration du champ de normales estimé n'est pas triviale. D'une part, elle repose sur l'intégrabilité de la surface, qui n'est pas garantie (cf. paragraphe 4). D'autre part, un biais dans l'estimation du champ de normales peut provoquer une déformation importante du relief. C'est pourquoi il a été suggéré d'utiliser la PS en complément d'autres techniques de reconstruction 3D qui estiment directement le relief, comme la stéréoscopie [15].

Il a déjà été dit que la PS procurait des cartes de profondeur sans trou. Cependant, puisque l'objet à reconstruire est photographié sous un seul angle, il est impossible d'effectuer une reconstruction 3D complète. Pour cela, il faudrait soit fusionner plusieurs cartes de profondeur acquises sous différents angles [4], soit là encore utiliser la PS en complément d'une autre technique de reconstruction 3D, comme par exemple le *shape-from-silhouettes* [9].

Outre ces difficultés chroniques de la PS, la mise en œuvre de cette technique repose sur un certain nombre d'hypothèses plus ou moins réalistes.

3.3 Remise en question des hypothèses

Bien que le *shape-from-shading* soit mal posé et que l'utilisation de $m \geq 3$ éclairages non coplanaires rende la PS bien posée, ces deux techniques de reconstruction 3D reposent fondamentalement sur les mêmes hypothèses [5].

Les trois hypothèses suivantes servent à ne pas avoir un trop grand nombre d'étalonnages à effectuer :

- Le récepteur photosensible est linéaire.
- Les aberrations de l'objectif sont négligées (distorsion radiale, vignettage, etc.).
- Le facteur $\cos^4 \alpha$ de l'équation de luminance (*image irradiance equation*, cf. [13]) est assimilé à 1.

Deux autres hypothèses semblent faciles à garantir en conditions de laboratoire :

- Les images sont nettes (pas de flou de mise au point).
- La surface de l'objet à reconstruire est opaque.

En revanche, l'hypothèse suivante, qui est difficile à justifier, vise surtout à simplifier le problème :

- Les réflexions secondaires de la lumière sur l'objet lui-même, ou sur d'autres objets voisins, sont ignorées.

Aucune de ces hypothèses ne sera remise en question ici.

Par ailleurs, la modélisation de la PS sous la forme $I = MS$ n'est possible qu'en faisant quatre autres hypothèses :

- L'éclairage consiste en un faisceau parallèle et uniforme.
- L'objet à reconstruire ne présente ni arête ni occultation, si bien que la normale N est définie partout.
- La surface vérifie la contrainte d'intégrabilité.
- La surface est lambertienne.

Se passer de la première de ces hypothèses reviendrait à une remise en question profonde du modèle, comme cela a d'ailleurs été fait pour le *shape-from-shading* [2].

Dans cet article, nous concentrons notre attention sur les trois dernières hypothèses. La deuxième et la troisième sont liées à la géométrie de l'objet photographié. On ne peut définir de normale qu'en un point où la *profondeur*, qui constitue une fonction des coordonnées de l'image, est de classe C^1 , ce qui n'est pas le cas le long d'une arête ou d'une occultation. D'autre part, la contrainte d'intégrabilité (10) n'est valide que si la profondeur est de classe C^2 . Par conséquent, un objet comportant des arêtes ou des occultations ne peut vérifier cette dernière. Quelle influence de tels points ont-ils sur la précision des résultats obtenus par stéréophotométrie ?

4 Contrainte d'intégrabilité

4.1 Effet d'une arête ou d'une occultation

En un point situé sur une arête ou une occultation, la normale n'est pas définie. Néanmoins, cette question ne se pose pas avec des images numériques, car un pixel p ne peut être entièrement non différentiable : tout au plus chevauche-t-il une ligne \mathcal{C} de non différentiabilité, qui coupe p en deux morceaux d'aires a^* et a^{**} (cf. figure 1). Bien entendu, les m niveaux de gris de p sont les moyennes pondérées des niveaux de gris correspondant aux deux vecteurs $M^* = \rho^* N^*$ et $M^{**} = \rho^{**} N^{**}$ qui caractérisent la surface de part et d'autre de \mathcal{C} :

$$I_p = \frac{a^*}{a^* + a^{**}} M^* S + \frac{a^{**}}{a^* + a^{**}} M^{**} S \quad (14)$$

Comme par ailleurs on postule que $I_p = \rho_p N_p S$, il vient :

$$\rho_p N_p = \frac{a^* \rho^* N^* + a^{**} \rho^{**} N^{**}}{a^* + a^{**}} \quad (15)$$

Dans le cas où $N^{**} = N^*$, on tire facilement de (15) :

$$\rho_p = \frac{a^* \rho^* + a^{**} \rho^{**}}{a^* + a^{**}} ; \quad N_p = N^* \quad (16)$$

Cela signifie que si p ne contient ni arête ni discontinuité, mais est constitué de deux couleurs différentes ρ^* et ρ^{**} , alors sa couleur ρ_p est une moyenne pondérée de ρ^* et ρ^{**} . En revanche, cela n'est plus vrai si le pixel est situé à cheval sur une arête ou une occultation. Par conséquent, en un tel pixel, l'albédo et la normale estimés par PS diffèrent des valeurs auxquelles on pourrait s'attendre, à savoir les moyennes pondérées suivantes :

$$\tilde{\rho}_p = \frac{a^* \rho^* + a^{**} \rho^{**}}{a^* + a^{**}} ; \quad \tilde{N}_p = \frac{a^* N^* + a^{**} N^{**}}{\|a^* N^* + a^{**} N^{**}\|} \quad (17)$$

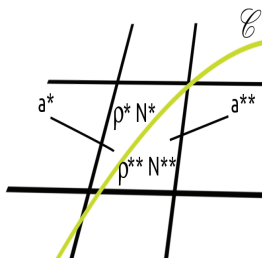


FIGURE 1 – La ligne \mathcal{C} , qui matérialise une arête ou le bord d'une occultation, coupe le pixel p en deux morceaux d'aires a^* et a^{**} , caractérisés par des vecteurs $M^* = \rho^* N^*$ et $M^{**} = \rho^{**} N^{**}$ qui peuvent être très différents.

4.2 Effet des écarts à l'intégrabilité

Nous venons de voir que l'estimation de l'albédo et de la normale par PS en un pixel situé sur une arête ou une occultation est erronée. Mais nous avons implicitement supposé,

dans le paragraphe 4.1, que la matrice S était connue. Or, il se trouve que l'estimation de cette matrice, dans le cadre de l'UPS, repose en particulier sur l'hypothèse que la surface est intégrable. On peut donc craindre que la présence d'arêtes ou d'occultations ne dégrade l'estimation de S . La figure 2-gauche montre en gros plan une image de synthèse (parmi huit) d'une théière texturée. L'anse de la théière provoque à la fois des occultations, sur ses bords gauche et droit, et des arêtes aux endroits où elle rejoint le corps de la théière. L'écart angulaire moyen sur les éclairages passe de $EAM = 14,36$ degrés à $EAM = 9,39$ degrés lorsqu'on retire de la matrice I les niveaux de gris des pixels où la contrainte d'intégrabilité est violée (cf. figure 2-droite).



FIGURE 2 – À gauche et au centre : une image de synthèse (parmi huit) d'une théière très texturée, lambertienne. L'écart angulaire moyen sur les éclairages vaut $EAM = 14,36$ degrés. À droite : détection des pixels où la contrainte d'intégrabilité est violée, avec un seuil $s = 0,05$ (cf. paragraphe 4.3). En retirant ces pixels de I , on obtient $EAM = 9,39$ degrés.

4.3 Détection des points de non intégrabilité

Pour savoir si un point vérifie le critère d'intégrabilité (10), on peut comparer son résidu à un seuil. C'est effectivement un des moyens qui permet d'intégrer un champ de normales en présence de discontinuités [6], mais il convient de distinguer la situation calibrée (CPS) de la situation non calibrée (UPS). En effet, la détection des points de non intégrabilité n'est utile en CPS que pour l'intégration des normales, alors que cela est également nécessaire, dans le cadre de l'UPS, pour estimer les matrices S et M . Il est donc intéressant de savoir détecter les points de non intégrabilité à l'aide d'un prétraitement des données, préalablement à la résolution de l'UPS. L'une des possibilités pour détecter ces points est d'utiliser un algorithme de détection de contours. Cependant, bien que cet algorithme puisse détecter l'ensemble des points de non intégrabilité, plusieurs problèmes subsistent : tout d'abord, un tel algorithme ne prend pas en compte la redondance des images stéréophotométriques. Ensuite, nous ne sommes pas assurés que l'ensemble des points détectés fasse partie d'un contour (un changement brutal de texture sera perçu comme un contour par ce genre d'algorithme), ce qui implique que nous pouvons perdre une certaine quantité d'information. La méthode que nous proposons permet de limiter la perte d'information, en utilisant le fait que nous avons affaire à une série d'images stéréophotométriques. Pour ce faire, nous comparons le vecteur I_p des m niveaux de gris d'un pixel

p aux vecteurs I_v des m niveaux de gris de ses huit plus proches voisins v , grâce à la distance suivante :

$$d(p,v) = \left\| \frac{I_p}{\|I_p\|} - \frac{I_v}{\|I_v\|} \right\|_\infty \quad (18)$$

En normalisant les niveaux de gris dans (18), on s'assure qu'une valeur élevée de $d(p,v)$ ne correspond pas à une variation d'albédo [11] et, en l'absence d'écarts au modèle lambertien (cf. paragraphe 5), il est raisonnable de conclure qu'un pixel p jouxte une arête ou une occultation si et seulement si la valeur maximale de la distance $d(p,v)$ correspondant aux huit plus proches voisins v dépasse un seuil s ($s = 0,05$ dans le cas de la figure 2).

5 Hypothèse lambertienne

5.1 Écarts à l'hypothèse lambertienne

La loi de Lambert permet d'obtenir une modélisation linéaire de la PS. Notons que cela n'est pas le cas du *shape-from-shading*. En effet, comme la normale N est normée, elle doit vérifier la contrainte non linéaire $\|N\| = 1$. Or, lorsqu'on dispose d'une seule image ($m = 1$), il n'est pas raisonnable de supposer l'albédo ρ inconnu, car cela rendrait le problème encore plus mal posé. Pour $m = 2$, il est encore nécessaire de supposer l'albédo connu et de résoudre un système de $m = 2$ équations non linéaires à 2 inconnues. C'est donc seulement lorsque $m \geq 3$ que le modèle de la PS devient linéaire.

Néanmoins, la linéarité de la PS sous l'hypothèse lambertienne est illusoire. D'une part, $M = \rho N$ n'est pas un vecteur quelconque de \mathbb{R}^3 , puisque $\|M\| = \rho \leq 1$. Par conséquent, comment estimer l'albédo et la normale en un point p où $\|\widehat{M}_p\| > 1$? Le même genre de dilemme se pose lorsque le vecteur \widehat{M}_p forme avec la direction d'observation un angle supérieur à $\pi/2$, ce qui ne devrait jamais se produire avec un objet opaque.

D'autre part, la loi (1) n'est valide que si le produit scalaire $N_p S^i$ est positif, car un niveau de gris est proportionnel à une énergie lumineuse, qui est forcément positive. Il faut donc remplacer la loi (1) par :

$$I_p^i = \max\{\rho_p N_p S^i, 0\} \quad (19)$$

Les points tels que $N_p S^i < 0$ constituent les *ombres propres*. En l'absence d'arêtes et d'occultations, la limite entre une partie éclairée et une ombre propre est constituée de pixels où $N_p S^i = 0$ et s'appelle un *terminateur*. La prise en compte des ombres propres dans (19) fait perdre à la loi de Lambert son caractère linéaire, sur laquelle repose en grande partie la résolution « classique » des problèmes CPS et UPS (cf. paragraphe 2). Pour restaurer la linéarité, on peut soit modifier les termes concernés de la matrice I , en acceptant qu'ils deviennent négatifs (cf. paragraphe 5.3), soit les retirer purement et simplement de I (cf. paragraphe 5.4). Notons que Hertzmann et al. ont montré dans [11] comment s'accommoder de la non linéarité de

la loi (19), en intégrant à la scène une sphère recouverte du même matériau que l'objet à reconstruire, ce qui permet d'apprendre la relation entre normales et niveaux de gris, y compris d'ailleurs pour un matériau non lambertien.

Un deuxième type d'écarts à la loi de Lambert linéaire (1) provient des *ombres portées*. Contrairement aux ombres propres, les ombres portées ne peuvent pas être caractérisées localement, donc l'approche de Hertzmann et al. ne peut y remédier.

Enfin, on est en droit de se demander si une loi aussi simple que la loi de Lambert peut modéliser la diffusion de la lumière sur des objets réels. Aussi surprenant que cela puisse paraître, les matériaux mats sont « plutôt bien » modélisés par la loi de Lambert non linéaire (19). Néanmoins, de nombreux autres modèles ont été proposés, parmi lesquels le modèle de Phong est un des plus souvent rencontrés [16]. Ce modèle rajoute à la loi de Lambert une « composante spéculaire » censée modéliser les reflets brillants :

$$I_p^i = \max\{\rho_p N_p S^i + \rho'_p (V R^i)^\alpha, 0\} \quad (20)$$

où V désigne la direction d'observation, R^i caractérise le faisceau lumineux S^i après réflexion sur la surface (selon la première loi de Descartes), α est un paramètre du modèle et ρ'_p constitue une nouvelle inconnue en p .

La figure 3 illustre ces trois types d'écarts au modèle lambertien linéaire (ombres propres, ombres portées et taches spéculaires) sur l'image de synthèse d'une théière. Chacun des écarts affecte une zone très localisée de la théière. Par conséquent, une grande partie des niveaux de gris des deux images du haut de la figure 3 sont identiques.



FIGURE 3 – À gauche : image simulée d'une théière, utilisant le modèle de Phong (20) et prenant en compte les ombres portées. Les trois types d'écarts au modèle lambertien sont présents sur cette image. En bas : valeur absolue des écarts (en niveaux de gris : blanc pour 1, noir pour 0) entre le modèle lambertien (1) et le modèle de Phong (20).

5.2 Effet des écarts à la loi de Lambert

On doit bien sûr s'attendre à ce que tout écart à la loi de Lambert linéaire dégrade les résultats obtenus par PS. Cela est illustré sur l'exemple de la figure 4, où l'on constate effectivement que les normales sont très mal estimées là où la loi de Lambert est mise en défaut. Les résultats seront a fortiori encore plus décevants dans le cadre de la stéréophotométrie non calibrée (UPS), où ces écarts se traduiront à la fois par une mauvaise estimation des éclairages et par une mauvaise estimation des normales.

Le plus sûr moyen de s’affranchir de l’influence des niveaux de gris qui ne suivent pas le modèle lambertien linéaire (1), qui constituent des « données aberrantes » (ou *outliers*), consiste à les retirer de la matrice I . C’est ce que nous ferons dans le paragraphe 5.4, mais nous allons commencer par explorer une autre voie.

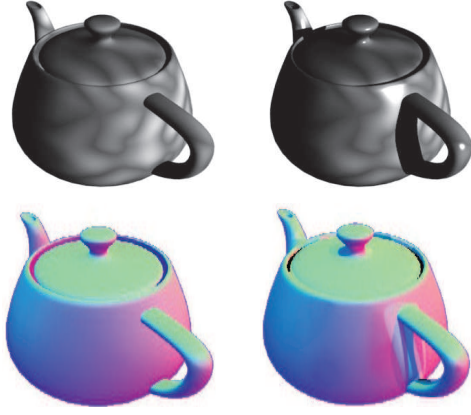


FIGURE 4 – Résolution de la CPS en présence d’écarts au modèle lambertien. À gauche, de haut en bas : une image de synthèse (parmi huit) d’une théière lambertienne sans ombre portée ; champ de normales estimé, codé dans les trois canaux (R,V,B). À droite, de haut en bas : idem lorsque la théière comporte des taches spéculaires et des ombres portées. Le résultat en bas à droite est très dégradé.

5.3 Correction des écarts à la loi de Lambert

Une première façon de s’affranchir de l’influence des données aberrantes consiste à les *corriger*, c’est-à-dire à concevoir un procédé permettant de rendre négatif le niveau de gris d’un pixel situé à l’intérieur d’une ombre propre et positif celui d’un pixel situé dans une ombre portée, et à modifier celui d’un pixel situé dans une tache spéculaire, afin de forcer tous les niveaux de gris à vérifier le modèle linéaire (1). Ce procédé présente l’avantage de n’écarter aucun pixel, et donc de pouvoir estimer par la suite l’albédo et la normale en tout point de l’objet, même là où des niveaux de gris initiaux sont manifestement erronés.

Pour effectuer une telle correction, le plus simple est de s’inspirer de la résolution du problème UPS, qui consiste à effectuer la décomposition SVD de la matrice I puis à la remplacer par la matrice \bar{I} définie par (7), qui est de rang 3. En pratique, cette correction n’est valide que si le pourcentage de données aberrantes reste faible, car ce nombre influe fortement sur les trois plus grandes valeurs singulières de I . Un exemple de correction des données obtenu avec cette méthode est présenté sur la ligne du milieu de la figure 5 : à gauche, les valeurs négatives de \bar{I} sont mises à 0 ; à droite, une constante a été rajoutée à l’ensemble des niveaux de gris, afin de mettre en évidence les valeurs négatives obtenues à l’intérieur des ombres propres.

Wu et al., qui s’inspirent de travaux présentés dans [14],

recommandent plutôt de décomposer la matrice I comme la somme d’une matrice A « approximativement de rang 3 » et d’une matrice d’erreur E la plus creuse possible [20], ce qui revient à résoudre le problème d’optimisation :

$$\min_{(A,E)} \{\text{rang}(A) + \lambda \|E\|_0\} \quad (21)$$

où $\|E\|_0$ est la norme L^0 (égale au nombre de termes non nuls de la matrice E), qu’on cherche à minimiser pour rendre E creuse, et $\lambda > 0$ est un hyper-paramètre. Wu et al. montrent que le problème (21) admet « à peu près » la même solution que :

$$\min_{(A,E)} \{\|A\|_* + \lambda \|E\|_1\} \quad (22)$$

où $\|A\|_*$ est égal à la somme des valeurs singulières de A , et $\|E\|_1$ est la norme L^1 . Ils proposent une méthode de résolution efficace (appelée ALM) du problème (22). La matrice A peut ensuite être considérée comme une version corrigée de la matrice I des niveaux de gris.

Au regard des corrections obtenues, cette méthode surpasse la précédente, mais seulement si λ est convenablement choisi. Il semble donc que λ constitue un paramètre « critique ». En outre, et comme le montre la figure 5 (en haut à droite), la correction des ombres portées est décevante. Ceci montre que la correction des images, même si elle constitue une piste intéressante, ne permet guère de prétraiter un grand nombre de données aberrantes. Dans le cadre du problème UPS, il nous semble plus pertinent de détecter les données aberrantes, pour pouvoir les écarter.

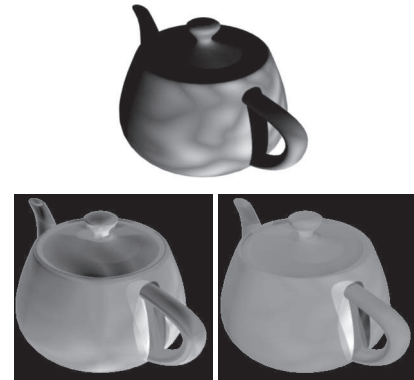


FIGURE 5 – En haut : image de synthèse présentant des écarts à la loi de Lambert. En bas : écarts entre l’image d’origine et les images corrigées par la méthode ALM (à gauche) et par utilisation d’une SVD (à droite).

5.4 Détection des écarts à la loi de Lambert

Concentrons-nous dorénavant sur le problème UPS, qui a trois objectifs : estimer l’albédo, le champ de normales et les éclairages. L’estimation de l’albédo et des normales pour les pixels situés à l’intérieur d’une ombre ou d’une tache spéculaire semble délicate, mais l’estimation des

éclairages ne nécessite pas d'utiliser l'ensemble des pixels, puisque théoriquement $3m$ pixels suffisent pour estimer les m vecteurs S^i . Plutôt que de chercher à corriger les données aberrantes, il semble donc plus judicieux de les écarter, afin d'utiliser seulement les pixels qui sont conformes à la loi de Lambert linéaire (1) dans les m images.

Pour ce faire, plusieurs approches peuvent être envisagées. Un pixel situé dans une ombre (propre ou portée) étant caractérisé par un niveau de gris faible, et un niveau de gris élevé indiquant au contraire que le pixel se situe dans une tache spéculaire, la technique la plus simple consiste à seuiller les différentes images [3]. Néanmoins, de tels seuils sont difficiles à fixer, et si le nombre m d'images devient trop important, il est possible que tout pixel soit écarté à cause d'un de ses m niveaux de gris. On risque ainsi d'écarter la totalité des pixels !

Nous proposons plutôt de mesurer la distance euclidienne $d(p) = \|I_p - \bar{I}_p\|$ du vecteur I_p des niveaux de gris d'un pixel p au vecteur \bar{I}_p de ses niveaux de gris corrigés (cf. paragraphe 5.3), obtenus soit par la SVD (7) soit par la méthode ALM [20]. On peut ainsi trier les pixels par ordre croissant, selon cette distance, et n'utiliser que les $q\%$ premiers pixels ainsi triés. Si les images ne présentent pas « trop » de défauts, il est probable qu'une distance $d(p)$ élevée correspond à un écart au modèle dans une image au moins (cf. figure 7-à droite).

Il reste cependant à fixer le pourcentage q de pixels à conserver. Pour cela, il faut trouver un compromis entre la mise à l'écart des données aberrantes et une certaine stabilité numérique (cf. figure 6). Puisque plus m est grand, plus la probabilité qu'un pixel présente un défaut dans une image est élevée, nous proposons le choix $q = \exp(-\lambda(m - 3))$. Dans la suite nous considérons $\lambda = 0,1$.

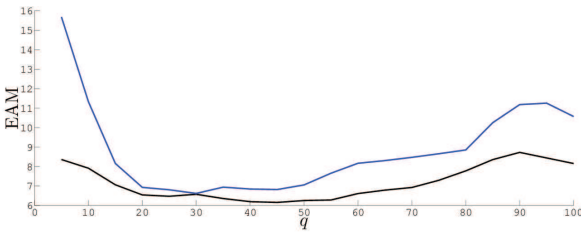


FIGURE 6 – Évolution de l'EAM sur les éclairages, en fonction du pourcentage q de pixels les plus proches de la projection SVD (en noir), ou de la correction ALM (en bleu). On a ici $m = 12$ images de test : 'horse', disponibles sur le web.

Rappelons enfin que le pourcentage de données aberrantes influe fortement sur les valeurs singulières de la matrice I , et donc sur les distances $d(p) = \|I_p - \bar{I}_p\|$. La figure 7 montre comment la largeur des taches spéculaires, que l'on peut contrôler par le paramètre α (la tache grossit lorsque α diminue), influe sur cette méthode de détection des données aberrantes, qui offre toutefois un niveau de robustesse raisonnable.

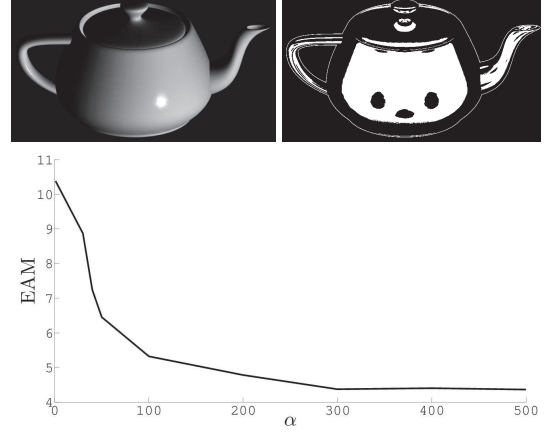


FIGURE 7 – En haut : une (parmi huit) image de synthèse calculée selon le modèle de Phong avec $\alpha = 500$ (à gauche), et détection des données aberrantes (à droite, en noir). En bas : évolution de l'EAM sur les éclairages, en fonction de α (cf. equation (21)), en détectant les données aberrantes (ici $m = 8$ et $q \simeq 61\%$).

6 Tests

Nous présentons enfin quelques résultats relatifs à la résolution du problème UPS, afin de montrer l'intérêt de détecter les données aberrantes selon la méthode décrite dans le paragraphe 5.4. Les données (réelles) testées sont $m = 8$ images pour chacun des 7 premiers jeux d'images de visages provenant de la base Yale DataFaceB [8], pour lesquelles une estimation précise des éclairages est fournie.

En mesurant le temps de calcul et l'EAM sur les éclairages, nous comparons la méthode proposée aux deux méthodes de correction décrites dans le paragraphe 5.3 : correction par SVD et méthode ALM [20]. Nous choisissons d'évaluer également RANSAC [7], qui constitue un estimateur robuste aux données aberrantes que sont les trois types d'écarts à la loi de Lambert. Cet algorithme nécessite de fixer l'échelle du bruit, là où d'autres estimateurs imposent un pourcentage maximal de données aberrantes [18] : moindre médiane des carrés (*Least Median of Squares*) ou moindres carrés tronqués (*Least Trimmed Squares*).

Les résultats obtenus en utilisant $m = 8$ images de chaque visage sont consignés dans le tableau 1. La simplicité de la méthode proposée la rend sensiblement plus rapide que les trois autres méthodes testées, pour une efficacité au moins comparable. D'autre part, l'évolution de l'EAM sur les éclairages, en fonction du nombre m d'images utilisées, est représentée sur la figure 8.

7 Conclusion et perspectives

Dans cet article, nous avons passé en revue les différentes hypothèses classiques de la stéréophotométrie. Nous suggérons des pistes pour nous affranchir notamment de deux d'entre elles : la *contrainte d'intégrabilité* et l'*hypothèse*

Données	YaleB01_P00	YaleB02_P00	YaleB03_P00	YaleB04_P00	YaleB05_P00	YaleB06_P00	YaleB07_P00
SVD	15,60 (0,56 s)	8,56 (0,55 s)	17,38 (0,56 s)	8,38 (0,53 s)	7,44 (0,61 s)	14,31 (0,54 s)	8,43 (0,54 s)
ALM	6,54 (0,62 s)	4,48 (0,61 s)	6,08 (0,65 s)	7,50 (0,63 s)	9,08 (0,83 s)	7,88 (0,66 s)	2,40 (0,63 s)
RANSAC	19,02 (34,44 s)	7,18 (32,38 s)	16,51 (29,41 s)	5,90 (45,64 s)	5,83 (34,35 s)	8,01 (42,30 s)	10,43 (37,35 s)
Notre méthode	4,50 (0,17 s)	5,76 (0,15 s)	4,05 (0,17 s)	3,10 (0,18 s)	6,07 (0,23 s)	4,99 (0,18 s)	4,72 (0,17 s)

TABLE 1 – Comparaison de notre méthode à la correction par SVD (utilisant la totalité du masque) ou par ALM (en sélectionnant les pixels les plus proches de la correction) [20], et à la méthode RANSAC. Après sélection des points du masque à utiliser, nous estimons les éclairages [17]. Les résultats sont l'EAM (en degrés) sur les éclairages et le temps de calcul (en secondes), pour $m = 8$ images, provenant de la base de données de Yale.

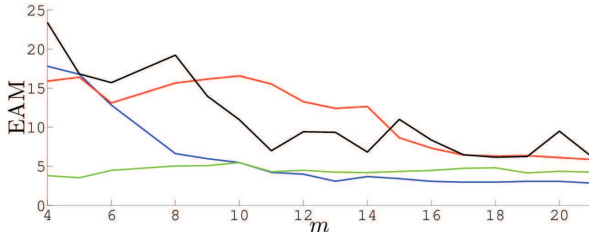


FIGURE 8 – Évolution de l'EAM sur les éclairages, en fonction du nombre m d'images utilisées. En rouge : remplacement de I par la matrice \bar{I} la plus proche de rang 3. En bleu : méthode ALM [14]. En noir : RANSAC. En vert : notre méthode. Images de test : YaleB01_P00.

lambertienne, qui sont les hypothèses fondamentales de la résolution du problème de la stéréophotométrie non calibrée (UPS). Nous proposons en particulier une méthode simple et efficace de détection des données aberrantes que sont les ombres (propres ou portées) et les reflets spéculaires, et nous prouvons expérimentalement comment la détection de ces écarts permet effectivement d'améliorer l'estimation des éclairages du problème UPS.

Avant de remettre en question un certain nombre d'autres hypothèses, nous devons prioritairement venir à bout de l'estimation de l'albédo et des normales pour les pixels assimilés à des données aberrantes. Il conviendra ensuite d'adapter la résolution de l'UPS à des modèles de BRDF plus complexes. Cela sera facilité par notre méthode de détection des points aberrants, qui permet une estimation plus précise des éclairages.

Références

- [1] P. N. Belhumeur, D. J. Kriegman, et A. L. Yuille. The Bas-Relief Ambiguity. *IJCV*, 35(1):33–44, 1999.
- [2] M. Breuss, E. Cristiani, J.-D. Durou, M. Falcone, et O. Vogel. Perspective Shape from Shading: Ambiguity Analysis and Numerical Approximations. *SIAM Imaging Sciences*, 5(1):311–342, 2012.
- [3] B. Bringier, A. Bony, et M. Khoudair. Specularity and shadow detection for the multisource photometric reconstruction of a textured surface. *JOSA*, 29(1), 2012.
- [4] J. Chang, K. Lee, et S. Lee. Multiview normal field integration using level set methods. In *CVPR 2007*, pages 1–8.
- [5] P. Daniel. *Peut-on extraire le relief d'une seule image ?* thèse, Université Paul Sabatier, Toulouse, 2000.
- [6] J.-D. Durou, J.-F. Aujol, et F. Courteille. Integration of a Normal Field in the Presence of Discontinuities. In *EMMCVPR 2009, LNCS 5681*, pages 261–273.
- [7] M. Fischler et R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [8] A. S. Georgiades, D. J. Kriegman, et P. N. Belhumeur. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *PAMI*, 23(6):643–660, 2001.
- [9] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, et R. Cipolla. Multiview Photometric Stereo. *PAMI*, 30(3):548–554, 2008.
- [10] C. Hernández, G. Vogiatzis, B. Stenger, et R. Cipolla. Video normals from colored lights. *PAMI*, 33(10):2104–2114, 2011.
- [11] A. Hertzmann et S. M. Seitz. Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *PAMI*, 27(8):1254–1264, 2005.
- [12] B. K. P. Horn. Understanding Image Intensities. *Artificial Intelligence*, 8(2):201–231, 1977.
- [13] B. K. P. Horn. Obtaining Shape from Shading Information. In *Shape from Shading*, pages 123–171. MIT Press, 1989.
- [14] Z. Lin, M. Chen, et Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices. Tech. Rep. UILU-ENG-09-2215, UIUC, 2009.
- [15] D. Nehab, S. Rusinkiewicz, J. Davis, et R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. In *SIGGRAPH 2005*, pages 536–543.
- [16] B. T. Phong. Illumination for computer generated pictures. In *Communications of ACM 1975*, pages 311–317.
- [17] Y. Quéau et J.-D. Durou. Résolution du problème de la stéréophotométrie non calibrée par estimation de l'intensité des éclairages. In *Orasis 2013 (soumis)*.
- [18] P. J. Rousseeuw et A. M. Leroy. *Robust regression and outlier detection*. John Wiley and Sons, 1987.
- [19] R. J. Woodham. Photometric Method for Determining Surface Orientation from Multiple Images. *Opt. Engin.*, 19(1):139–144, 1980.
- [20] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, et Y. Ma. Robust Photometric Stereo via Low-Rank Matrix Completion and Recovery. In *ACCV 2010, LNCS 6494*, pages 703–717.
- [21] A. L. Yuille et D. Snow. Shape and Albedo from Multiple Images using Integrability. In *CVPR 1997*, pages 158–164.